

# 분류전문가시스템에 관한 연구

김 정 현\*

## < 목 차 >

- |                              |                  |
|------------------------------|------------------|
| I. 서 론                       | 2. 분류전문가 시스템의 사례 |
| II. 전문가 시스템의 개요              | IV. 결 론          |
| 1. 전문가시스템의 개념 및 특성           | 참고문헌             |
| 2. 전문가시스템의 구성                | 영문초록             |
| III. 분류전문가시스템의 개발            |                  |
| 1. 분류업무와 자동화, 그리고<br>전문가 시스템 |                  |

## I. 서 론

정보화사회를 맞이하여 최근 각종 정보와 문헌이 급증하고 있으며, 아울러 이용자의 정보요구도 더욱 다양하고 신속성을 요하고 있다. 그런데 도서관업무의 대부분이 자동화됨에 따라 이용자에 대한 정보봉사는 과거 어느때 보다도 신속 정확하게 이루어지고 있지만 분류업무만은 아직도 전통적인 수작업방법에 의존하고 있어 도서관에서 자료정리에 필요한 시간과 노력은 그만큼 가중되고 있다고 하겠다. 그결과 이용자에게 필요로 하는 자료가 제때에 정리되지 않고 지연되는 경우가 종종 발생하고 있으며, 또한 문헌량의 증가는 이용자가 원하는 정보를 효율적으로 검색하기가 더욱 곤란하게 되어 가고 있다.

이러한 문제의 해결책의 하나로 1960년대 부터 자동분류에 관한 연구가 이루어져 왔지만 이는 서가상의 배열을 위한 분류기호의 자동생성이 아니라 주로 정

\* 계명대학교 강사

보검색적인 측면에서 용어분류에 치우쳐 왔다. 그후 1980년대에 들어오면서 인공지능분야의 하나인 전문가시스템이 각종 전문적·경험적 지식분야에 응용되기 시작하면서 도서관에 있어 특히 지적인 판단이 요구되는 분류업무에 적용을 위해서도 활발하게 연구가 이루어 지고 있다.

그러나 분류전문가시스템의 설계에 있어서 가장 핵심이 되고 있는 지식베이스의 구축 및 규칙생성을 위한 전문가의 전문지식과 경험을 종합하여 규칙화하는 문제는 문헌분류업무를 시스템화 하는데 있어서 아직은 정형화되지 못하고 있는 부분이다. 또한 이미 시도된 연구도 일반적이며, 단편적이고 초보적인 연구에 그치고 있으므로 실제업무에는 거의 활용이 되지 않고 있는 실정이다.

본연구의 목적은 전문가시스템의 개념,특성 및 구성요소 등에 대하여 개괄적으로 살펴본 다음 이를 바탕으로 실제 문헌분류업무에 개발되어 있는 전문가시스템의 사례들을 분석하므로써 이들에 대한 문제점과 앞으로의 적용방안을 제시하는데 있다.

## II. 전문가시스템의 개요

### 1. 전문가시스템의 개념 및 특성

#### 1) 전문가시스템의 개념

전문가시스템에 대한 하나의 규정된 정의는 없지만 다음과 같은 여러사람들의 주장을 종합하여 보므로써 대체적인 개념을 알 수 있다.

먼저 Waterman 등은 "전문가시스템이란 어떤 분야에서의 여러 전문가의 지식을 구체화하므로써 시스템 이용자들이 유용한 추론(inference)을 할 수 있게 한 컴퓨터 프로그램"<sup>1)</sup> 이라고 하였으며, Bobrow 등은 "문제해결과정에서의 경험적

1) Donald A. Waterman and Fredrick Hayes-Roth. "An Investigation of Tools for Building Expert Systems," in *Building Expert Systems* / edited by Fredrick Hayes-Roth etc.(London : Addison-Wesley , 1983), p.169.

인 지식을 특정한 영역에 충분히 적용한 컴퓨터 프로그램”<sup>2)</sup> 이라고 하였다. Luconi 등은 “전문가시스템은 어려운 문제를 잘 해결하기 위하여 전문화된 기호적 추리력(reasoning)을 이용하는 컴퓨터 프로그램이다. 바꾸어 말하면 전문가시스템은 ①특정 문제분야에 대한 전문화된 지식을 이용한다. ②수치적 계산보다는 기호적이거나 질적인 추리력을 사용한다. ③비전문가인 인간보다는 나은 능력의 수준에서 문제를 해결한다”<sup>3)</sup> 고 정의하였다.

이상의 전문가시스템에 대한 정의에서 살펴본 바와 같이 인간전문가의 문제해결 방식을 모형화한 프로그램이 전문가시스템이며, 이 시스템을 구축하는 사람을 지식엔지니어(knowledge engineer)라 하는 바, 이들은 ①인간전문가를 연구하여 전문가가 이용하고 있는 사실과 경험의 법칙을 파악한 다음, ②실제 문제 해결 상황에서 이 전문가들이 사용하는 추론전략을 결정하므로써, ③인간전문가들의 행동을 시뮬레이트하는 시스템을 개발하게 된다.

## 2) 전문가시스템의 특성

전문가시스템과 전통적인 프로그램방식간의 차이는 전자는 지식, 후자는 데이터를 처리하는 것이라 할 수 있다. 하나는 지식을, 또 하나는 데이터를 처리하는 차이가 두 시스템으로 하여금 서로 다른 영역으로 갈라지게 하였다고 볼 수 있다. 물론 두 시스템이 컴퓨터와 프로그래밍 언어를 사용하여 정보를 처리한다는 점에 있어서 같은 영역으로 해석할 수 있겠지만 실제적으로 두 시스템간의 유사점보다는 차이점이 훨씬 많을 뿐만 아니라 시스템의 목적, 사용되는 용어, 시스템 개발에 참여하는 사람들의 기능과 역할, 컴퓨터의 이용정도 등에 있어서 뚜렷하게 구분된다.

Waterman은 두 시스템의 구체적인 차이를 <표1><sup>4)</sup>과 같이 구분하였다.

- 
- 2) D. G. Bobrow, S. Mittal, M.j. Stefik. "Expert Systems : Perils and Promise," *Communications of the ACM*, Vol.29, No.9(1986), pp.880-894.
  - 3) Fred L. Luconi etc. "Expert Systems; the Next Challenge for Managers," *Sloan Management Review*, Vol.27, No.4(Summer 1986), p.4.
  - 4) Donald A. Waterman. *A Guide to Expert Systems*. London : Addison-Wesley, 1986. p.24.

&lt; 표 1 &gt; 전통적인 프로그래밍과 전문가시스템의 비교

구 분	전통적인 프로그래밍	전문가시스템
解의 도출방식	알고리즘	휴리스틱
파일의 저장	테이타베이스	작업 메모리
데이타의 저장방식	수치데이타	기호로 된 지식베이스
데이타의 처리	수치지향적	기호지향적
처리방식	연속적 또는 일괄적	고도의 交互의 처리
중간단계의 설명력	처리 중간단계의 설명 불가능	처리 중간단계의 설명이 용이

한편 Rolston에 의하면 이상적인 전문가시스템은 전통적인 시스템에 비하여 ①해당 전문영역의 지식, ②탐색기법의 응용, ③발견적 분석의 지원, ④기존의 지식으로 부터 새로운 지식을 추론하는 능력, ⑤기호적 처리능력, ⑥자신의 사고를 설명하는 능력 등을 갖고 있다<sup>5)</sup> 했다.

결국 전문가시스템은 풍부한 지식을 가지고 좀 더 지능적인 시스템을 목표로 하며, 추론과정과 지식을 이용하여 이용자에게 의사결정을 내릴 수 있는 정보를 제공하는 반면에 전통적인 프로그래밍방식은 다양한 데이터구조와 알고리즘을 통해 이용자에게 사실적인 정보를 제공하는 것이라 할 수 있다.

## 2. 전문가시스템의 구성

전문가시스템은 일반적으로 <그림1><sup>6)</sup>과 같은 구조를 가지고 있다. 지식을 조직적으로 저장해 두는 지식베이스(knowledge)와 주어진 문제와 관련되는 지식들을 찾아서 문제를 해결해 나가는 효율적인 방법을 저장하는 추론엔진(inference engine)을 중심으로 지식획득시스템(knowledge acquisition facility)과 사용자 인터페이스(user interface)와 설명기구(explanation facility) 등으로 구성된다.

5) D.W.Rolston. *Principles of Artificial Intelligences and Expert Systems Development*. New York : McGraw-Hill, 1986. p.262.

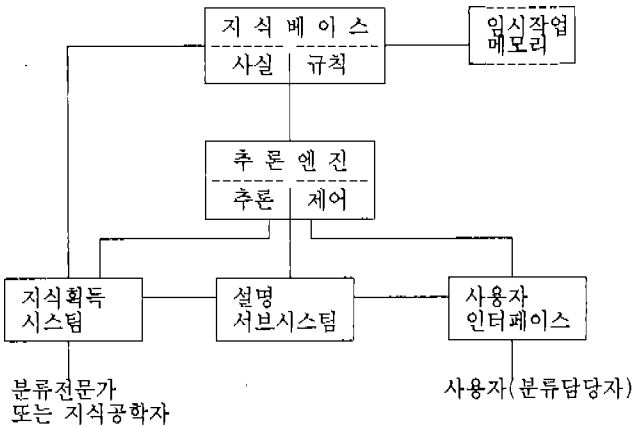
6) Paul Harmon and David King. *Expert Systems ; Artificial Intelligence in Business*. New York : John Wiley, 1985. p.49

지식베이스에는 사실(fact)과 규칙(rule)이 저장되는데 사실은 데이터나 사물에 대한 묘사를 말하며, 규칙은 어떤 결정을 내리기 위해 사실을 이용하는 범칙을 말한다.

추론엔진은 새로운 지식을 추론하기 위해 규칙들을 어떻게 적용해야 할 것인가를 결정하는 규칙해석기(interpreter)와 규칙들의 적용순서를 결정하는 스케줄러(scheduler)로 구성된다.

지식획득시스템은 전문가의 지식을 지식데이터베이스에 저장하는 기능을 갖고 있으며, 단순한 입력기능을 가질 수도 있지만 기계학습에 의한 규칙이나 사실을 저장할 수도 있다.

사용자 인터페이스는 사용자로부터 자료를 받아 들이고 결과를 알려 주는 역할을 하게 되는데 주로 메뉴방식이나 대화방식을 이용하고 있다. 사용자 인터페이스의 보조시스템으로써 설명시스템을 가지고 있어야 한다. 이는 시스템의 지식을 사용자가 이해할 수 있도록 하고 처리과정을 설명해 주어야 하기 때문이다.



< 그림 1 > 전문가시스템의 구성

### Ⅲ. 분류전문가시스템의 개발

#### 1. 분류업무와 자동화, 그리고 전문가시스템

분류업무란 일반적으로 정보자료를 그의 주제에 따라서 배열하고 형식에 따라서 구분하기 위한 체계적인 조직 및 그 조직에 따라서 정보자료를 해당위치에 배정하는 것인데<sup>7)</sup> 다음과 같은 과정을 거쳐 분류기호를 결정하게 된다.

- ① 분류대상인 도서의 주제를 분석하여 파악한다.
- ② 주제를 구성하고 있다고 생각되는 중요개념을 추출한다.
- ③ 분류표의 사용지침이나 분류규칙에 따라 상술한 주제구성개념에 대응하는 분류기호를 분류표에서 선택하여 부여한다. 이 과정에서 상관색인을 이용하여 중요개념에 해당하는 분류기호를 선택하는 경우가 많다.

그런데 이러한 분류행위의 결과인 분류기호는 채택하고 있는 분류체계내에서 특정주제를 확인하고, 서가상 혹은 목록기입상 자료를 명확하게 배열할 수 있도록 해주는 도구가 될 뿐만 아니라 서지데이터베이스의 검색수단으로 이용되어 검색시스템의 적합율과 재현율을 높이는 데에도 활용될 수 있다.

따라서 분류기호는 체계적이고 정확하게 부여하지 않으면 안된다. 분류자가 동일주제 자료에 대한 분류기호를 일관성없이 부여하게 되면 동일주제의 자료를 효과적으로 탐색하여 이용할 수 없게 되어 분류기호 본연의 기능을 상실하게 된다. 채택하고 있는 분류체계의 분류규칙과 분류에 필요한 전문지식을 일관성있게 적용시키는 것 만이 이러한 혼란을 막는 방법이 될 수 있다.

이러한 문제의 해결책의 하나로 시작된 연구가 자동분류라고 할 수 있으며, 이는 분류과정을 몇가지 알고리즘을 사용하으로써 기계적으로 처리하는 것을 의미하는데 1961년 Bayes' Theorem을 이용한 Maron의 연구<sup>8)</sup>에서 비롯되고 있

7) 정필모. 文獻分類論. 서울 : 구미무역, 1991. p.15.

8) M.E.Marón. "Automatic Indexing ; an Experimental Inquiry," *Journal of Association for Computing Machinery*, Vol.8, No.3(1961), pp.404-417.

다. 이후 각종 통계적 기법을 이용한 자동분류에 대한 연구가 수없이 이루어져 왔는데 크게 두가지 유형으로 구분할 수 있다. 즉, 특정한 주제에 관한 분류항목과 이에 대응하는 주제어들을 단서로 하여 문헌을 특정항목에 분류하는 유형과 문헌들이 갖는 주제어의 유사도에 따라 같은 집단으로 분류하는 문헌집단화(클러스터링)로 구분되는데 모두가 주제어에 기초하여 문헌과 분류항목, 또는 문헌과 문헌간의 유사도를 측정하는 과정에 의해 분류하였다.<sup>9)</sup> 이와같이 지금까지의 자동분류에 관한 연구는 주로 분류표에 의거하여 분류기호를 자동적으로 생성시키고자 하는 당초의 연구목적에서 벗어나, 컴퓨터에 의한 탐색의 효율성을 향상시키기 위하여 기계가독형 데이터베이스를 위해 적절히 그룹핑하고, 배열하고자 하는 문헌분류나 용어분류에만 치우쳐 왔다. 따라서 자동분류에 대한 지금까지의 연구는 주로 도서의 서가상의 배가를 위한 분류적인 측면보다는 검색적인 측면에서 이루어져 왔다.

한편 1980년대에 들어 오면서 인공지능분야의 하나인 전문가시스템에 대한 연구가 활발히 이루어 지면서 의료진단이나 분류업무에도 적용되기 시작하였는데 그 중에서 유명한 것이 스펀포드 대학교에서 개발한 의료진단용 시스템인 MYCIN이라고 할 수 있다. 의료진단에 있어서는 먼저 환자의 특성이나 병의 증상 등을 분류하는 과정이 우선되어야 한다. MYCIN의 경우처럼 분류를 위한 지식이 명확하고도 적절한 형태로 표현되어 이용될 수 있다면 기본적으로 도서관에서의 문헌분류작업에서도 전문가시스템을 응용할 수 있는 것이다. 1983년 Clarke와 Cronin<sup>10)</sup> 처음으로 그 적용 가능성을 시사한 이래 여러사람들에 의해 실험적인 연구가 이루어져 오고 있다. 이것은 분류표와 분류규정의 지식베이스화를 전제로 하며, 인간이 수행하는 분류작업의 수순을 정밀히 분석하여 그 과정을 가능한 한 정확하게 컴퓨터로 대체할 수 있도록 하는 것이다. 따라서 이는

9) 정영미. "도서관 정보학에서의 인공지능의 응용에 관한 고찰," 圖書館學, 제14집(1987), p.87.

10) A.Clarke and B.Cronin. "Expert Systems and Library/Information Work," *Journal of Librarianship*, Vol.15, No.4(1983), pp.277-292.

지식베이스를 어떻게 조직하고 분류자의 지적인 활동을 어떠한 방법으로 파악하느냐가 이것의 성패를 좌우한다.

종래의 분류는 경험을 통하여 학습되어 진다는 견해가 지배적이었으므로 그러한 과정을 분석하므로써 실제로 분류작업에 응용하려는 시도는 전무하였다고 해도 과언이 아니다. 도서관에서 분류업무편람이 거의 정비되어 있지 않은 이유 가운데 하나가 여기에 있다고 할 수 있다. 전문가시스템의 도입은 전통적인 도서관업무중에서 그다지 연구되지 않았던 분류과정에 새로운 전기를 가져오게 하므로써 보다 신속하고 효율적으로 일관성 있는 분류결과를 얻기 위한 수단으로서 유망하다고 하겠다.

## 2. 분류전문가시스템의 사례

앞서 언급한 바와 같이 1983년 Clarke와 Cronin<sup>11)</sup>이 UDC분류규칙에 대한 기본적인 지식표현규칙을 제시하면서 전문가시스템이 문헌을 분류하는데에도 응용될 수 있다고 처음으로 제안한 이래, 1985년 Burton<sup>12)</sup>은 DDC 19판을 대상으로 실제로 분류전문가시스템을 구현하였다. 이는 전문가시스템 개발도구인 ESP/Advisor와 Expert-Ease를 이용하였는데 실제 적용을 위해서는 여러가지 문제점이 있으며, 보완해야 할 사항들도 많다고 하겠다. 이후 여러 사람들에 의해 각종 분류표를 대상으로 다양한 방법으로 분류전문가 시스템이 개발되고 있는데 대표적인 사례를 중심으로 그 내용을 분석하여 보기로 한다.

### 1) ACN시스템

1988년 이시카와<sup>13)</sup>는 자동화된 도서관시스템에서 보조기능으로서의 인간과

11) *Loc. cit.*

12) P.F.Burton. "Expert System in Classification," in *Expert Systems in Libraries* / edited by F.Gibb(London : Taylon Graham, 1986), pp.50-66.

13) Ishikawa T. "The Man - Machine Interface Aspect of an Automatic Classification Numbering System in a Computerized Library System," *Journal of Information Processing*, Vol.11, No.3(1988), pp.199-205.  
石川徹也, 篠原まさ子. "日本十進分類法(NDC)を對象とする分類番號附與支援システム," 情報科學技術研究集會發表論文集, 第22回(1985), pp.201-209.

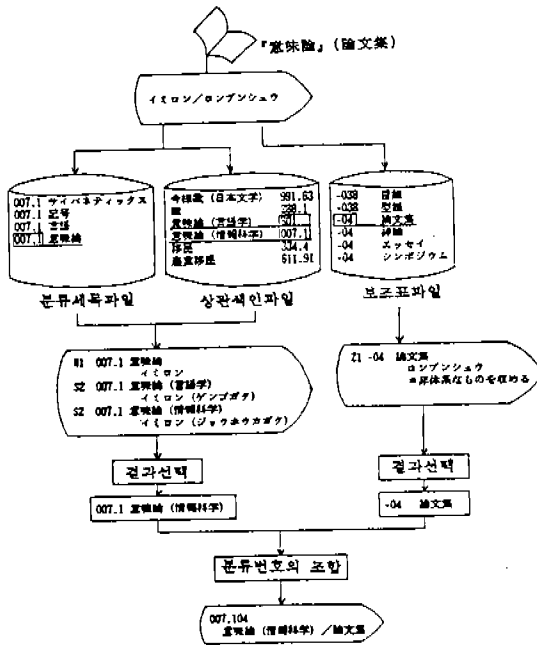


컴퓨터의 인터페이스 측면을 강조한 분류번호 자동부여(ACN)시스템을 개발하였는데 이는 두개의 서브시스템 즉, 일본십진분류법의 내용을 데이터베이스로 구축하여 이를 이용해 분류번호를 검색하는 분류번호 검색시스템과 일본십진분류법의 분류번호 합성규칙에 의해 최종적인 분류번호를 출력해 주는 분류번호 자동부여 서브시스템으로 구성되어 있다.

여기서 일본십진분류법의 내용을 담고 있는 데이터베이스가 바로 지식베이스라 할 수 있는데 이는 세계의 파일 즉, 분류제목파일과 상관색인파일, 그리고 보조표파일로 구성되어 있다. 검색시스템은 분석된 주제어, 주제분야, 자료의 형식등을 검색하여 사서가 선택할 수 있도록 되어 있다.

시스템을 이용할려면 사서는 우선 분류하고자 하는 자료의 주제분석을 하여 자료주제를 나타내는 용어(주제키)와 분야를 한정하는 용어(분야지시키) 및 자료형태를 나타내는 용어(형식분류키)를 분석, 추출하여야 한다. 그 다음 여기서 추출된 용어를 입력하여 시스템에서 분류제목파일, 상관색인파일, 보조표파일을 검색한 후 이것을 중간결과로 하여 파일에 있는 모든 정보를 표시한다. 사서는 이 중에서 분류해야 할 자료에 가장 적합하다고 생각되는 분류번호를 선택한다. 선택한 결과의 분류번호에 대하여 시스템은 자동적으로 조합하여 출력한다. <그림2>는 “의미론에 관한 논문집”이라는 도서에 대하여 분류번호를 부여하는 과정을 구체적으로 나타낸 것이다.

결국 이 시스템은 기계가독화된 일본십진분류표 데이터베이스를 기본전제로 삼고 있지만 개발 당시에는 임시로 필요한 주제분야만을 뽑아서 색인파일을 구성하였으며, 이용자위주의 대화형 시스템으로 인터페이스에 세심한 배려를 하였다는데 특징이 있다.



< 그림 2 > ACN시스템의 기능

## 2) NDC분류전문가시스템

1988년 호소노 등<sup>14)</sup>이 개발한 이 시스템은 ACN시스템과 마찬가지로 계층구조를 갖고 있는 NDC를 바탕으로 하고 있으며, “보험(339)”분야 도서를 대상으로 한 실험시스템이다. IBM에서 개발한 Expert System Environment(ESE)라는 전문가시스템 쉘을 이용하고 있으며, 지식은 생성규칙으로 표현되었다. 추론방식은 역방향추론을 사용하고 있다. 그리고 ESE에는 일본어가 지원되지 않으므로 키워드의 표기는 로마자이다. 한편 이 전문가시스템에서는 도서의 제목, 목차, 서문의 3개 정보원에서 인간이 추출한 키워드를 바탕으로 분류번호가 결정되며, 사용한 분류표는 계층구조를 갖고 있는 NDC이지만 NDC계층에는 논리적 모순

14) 細野公男, 田村俊作, 原田隆史. “圖書分類エキスパートシステム,” 情報學基礎, 8-4 (1988), pp.1-7.

이 있으므로 이를 바로 잡기 위해 모의NDC번호가 작성되어 있다. 따라서 NDC 분류번호의 참조는 이것을 경유하여 이루어지고 있다. 즉, 키워드를 모의 NDC 번호로 변환하는 규칙과 모의NDC번호를 다시 NDC번호로 변환하는 규칙에 의해 분류번호가 생성되는 것이다. 또 정보원의 차이에 따라 키워드의 영향력이 달라지도록 하기 위해 신뢰도의 값을 키워드의 출현장소에 따라 변화시키고 있다(제목은 0.8, 목차는 0.6, 서문은 0.5). 키워드가 복수의 정보원에 나타나는 경우는 대응하는 분류번호의 중요도를 높게 하고 계층상 같은 수준의 키워드가 다수 나타나는 경우는 각각의 분류번호 보다도 상위번호의 중요도가 높게 되도록 하고 있다.

이 시스템에 의해 분류담당자의 손으로 이미 분류된 보험분야 도서 41책을 대상으로 실험분류하여본 결과 31건밖에 일치하지 않았으며, 그 이유는 다음과 같다고 생각된다. ①키워드를 분류번호로 변환하는 규칙의 부족, ②키워드가 항상 정확하게 주제개념을 나타내지 않는 경우도 있다. ③모의NDC번호를 NDC번호로 변환하는 규칙의 불완전성, ④분류담당자와 시스템과의 사이에 분류의 수준(표현형식이나 지리구분에 대한 수준)에 있어 차이가 있다. 이러한 문제점들이 해결되므로써 이 시스템은 더욱 향상될 것이다.

### 3) UDC번호의 자동부여시스템

1988년 나카무라와 이시카와<sup>15)</sup>는 기계가독화된 국제십진분류표의 파일을 이용하여 임의의 개념에 대해 UDC번호의 자동부여를 위한 전문가시스템을 개발하였다. 국제십진분류표의 일본어 중간판을 컴퓨터조판 방식으로 인쇄하면서 만들어진 것이 PRISM시스템인데 이 시스템의 마스타파일로 부터 지식베이스의 바탕이 되는 분류표파일과 색인파일이 생성된 것이다. 이렇게 생성된 두가지의 기본파일과 개념파일, 보유파일, 그리고 이 파일들을 탐색하기 위한 여러가지 명령어 및 분류기호 결합규칙들로써 UDC의 분류번호 자동부여시스템이 구성되어

15) Nakamura Y. and Ishikawa T. "Expert System for Automatic UDC Numbering Assignment," *Proceedings of the 44th FID Conference and Congress part 1(Helsingki, 1988)*, pp.274-283.

있다.

이 시스템을 이용하여 적합한 UDC분류번호를 얻기 위해서는 여러 단계를 거치게 되는데 그 과정을 간단히 살펴 보면 먼저 분류대상의 자료내용을 분석하여 주요개념을 추출한다. 이 개념으로 색인파일이나 개념파일을 체크하여 분류번호를 탐색하며, 이때 해당개념과 일치하는 단어가 없으면 그와 동일한 또는 유사한 의미를 갖는 다른 형태의 단어들을 선택하거나 또는 시스템에게 조언을 구하면 시스템은 이용자를 메뉴로 안내하면서 대체 단어를 얻기 위한 여러가지 방법들을 제공한다. 만약 분류하고자 하는 자료의 내용이 복합개념으로 되어 있을 경우에는 먼저 파일을 체크하여 각 개념에 대한 최적 분류번호를 선택한 다음 임시메모리 영역에 저장한 후 이 분류번호들을 UDC의 분류번호 합성규칙에 의해 조합하여 최종적인 합성분류기호를 산출하게 된다.

이 시스템의 특징은 앞서 살펴 본 바와 같이 기계가독화된 분류표를 응용하여 지식베이스에 필요한 색인파일들을 자동생성하였다는데 있다.

#### 4) Herba Medica 시스템

1989년 Endres-Niggemeyer와 Schmidt<sup>16)</sup>는 ①실험용으로 간단한 예와 모의 시스템을 이용하여 패식분류표와 프레임표현간에 개념적 유사성을 설명하고, ②보다 포괄적인 시스템, 즉 분류전문가시스템의 개발을 위한 기초연구로서 인공지능과 분류연구, 정보학 등으로 부터 분류적인 지식처리에 대해 몇가지 접근방법을 소개한 논문을 발표하였다. 여기에서 실험용으로 개발한 분류전문가시스템이 Herba Medica 인데 이는 의약품과 관련문헌들을 분류할 수 있도록 모의 패식분류표를 미리 작성하여 Prolog시스템에서 프레임표현으로 구현되었다. <그림3>은 아편(opium)에 대한 프레임구조의 실례이다. 한편 Herba Medica 의 시스템 구성은 <그림4>와 같이 크게 지식베이스와 관리·지원시스템으로 구성되어 있다. 여기서 이 시스템의 가장 핵심이 되는 지식베이스는 유별로 계층화한 분류

16) Brigitte Endres-Niggemeyer and Bettina Schmidt. "Knowledge Based Classification Systems ; Basic Issues,a Toy System and Further Prospects," *International Classification*, Vol.16, No.3(1989), pp.146-156.

표와 약품 및 약품에 관련된 문헌 등으로 이루어져 있으며, 이 지식베이스의 주 위에는 이를 기동시키는 관리·지원시스템, 즉 사용자메뉴와 세개의 기능시스템 이 있다. 세개의 기능시스템은 ①자연언어로 적절한 클래스를 기술하면 분류코 드를 작성하고,역으로 분류코드는 클래스를 기술할 수 있는 classifier, ②분류표 를 검색하거나 갱신하는데 필요한 classification handler, ③지식베이스의 사실표 현부분으로 갱신 및 검색을 위한 database handler 등인데 이러한 것들은 이용 자를 위해 분류표를 브라우즈(browse)하고 필요에 따라 갱신하며, 시스템의 내 용을 설명해 준다는 것을 의미한다.

이 시스템의 특징은 지식베이스의 정형화를 위해서 패싯분류가 가장 적합할 것이라고 가정하고, 실제로 약품분야의 모의 패싯분류표를 만들어서 전문가시스 템을 구축하였다는데 있다. 그렇지만 다른 주제분야로 확장하기 위해서는 관련 지식의 패싯조직화가 선행되어야 할 것이다.

- Opium(taxo,value,Psomniferum.nil) - > ;
- Opium(teil,value,Frucht.nil) - > ;
- Opium(wirk,value,Morphin.Codein.Noscapin.Papaverin.nil) - > ;
- Opium(med,value,Analgetikum.nil) - > ;
- Opium(bem,value,"unterliegt"."Bet"."a."ubungsmittelgesetz"."nil) - > ;

< 그림 3 > Herba Medica의 지식베이스 프레임의 구조

user		menu
knowledge base		classifier (encoding, decoding)
classification scheme		classification handler (retrieval, update)
drugs	drug documents	database handler (retrieval, update)

< 그림 4 > Herba Medica 시스템의 구성

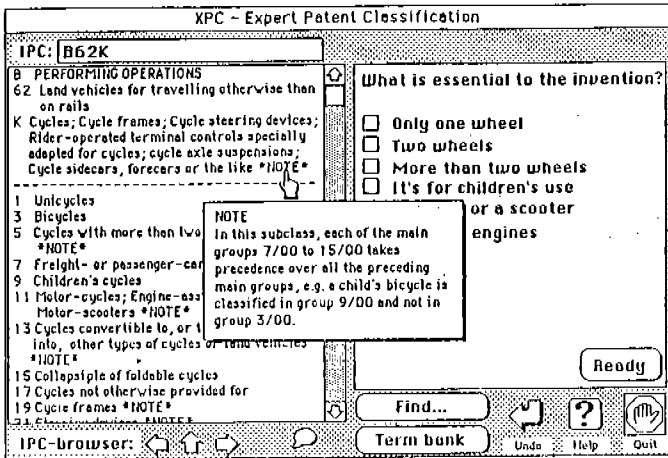
## 5) IPC전문가시스템

1991년 핀란드의 Valkonen과 Nykanen<sup>17)</sup>은 Maintosh PC와 Hypercard소프트웨어를 이용하여 국제특허분류(IPC)전문가시스템을 개발했다. 이는 분류법에 익숙하지 않은 이용자를 대상으로 한 모의시스템인데 특허분류의 기본적인 문제점을 제시하고 이용자 인터페이스에 의한 로컬언어로 접근하며, 시스템의 핵심이라고 할 수 있는 지식베이스와 추론엔진은 언어독립적인 점을 소개하고 있다.

일반적으로 특허분류는 기술내용을 기능별로 세분하는 방법과 용도별로 세분하는 방법이 있지만 IPC에서는 두가지를 병행한 기능지향형으로 되어 있다. 그리고 IPC의 구성은 모든 산업분야를 8개 부문(Section 이라고 함)으로 나누고, 그 아래 다시 하위계층으로 Class, Subclass, Group으로 구분하고 있다.

이와같이 IPC는 계층구조로 이루어져 있어서 자동화에 유리하다고 하겠다. <그림5>에서와 같이 화면의 왼쪽에 이용자가 선택한 IPC의 본문이 상단(상위주제)과 하단(하위주제)에 제시되며, 분류에 수반되는 주기와 참조사항은 키를 누르면 다시 화면에 나타난다. 이용자는 적절한 IPC코드를 선택하므로써 또 다른 참조내용을 볼 수 있다. 화면의 오른쪽에는 IPC본문을 지식베이스화한 시스템과 인터페이스할 수 있도록 되어 있으며,IPC 용어뱅크와 특수코드,그리고 기타 여러가지 기능들을 선택할 수 있다. 용어뱅크에는 IPC의 용어색인과 로컬언어를 포함한 일반언어가 포함되어 있다. 이 시스템은 핀란드의 많은 특허전문가들에게 소개되어 있으며, 비록 특허분류를 위해서 부족한 기능들이 많이 있지만 특허분류에 새로운 전기를 마련했다고 하겠다.

17) Pekka Valkonen and Olli Nykanen. "An Expert System for Patent Classification," *World Patent Information*, Vol.13, No.3(1991), pp.143-148.



< 그림 5 > IPC전문가시스템의 화면내용

### 6) Stratclyde 프로젝트

1988년 Sharif<sup>18)</sup>은 분류전문가시스템 개발의 타당성을 분석하는 프로젝트를 수행하였는데 이 프로젝트에서 그는 마이크로컴퓨터상에서 운영될 수 있는 세가지 종류의 셸, 다시말해 Expert-Ease, ESP/Advisor와 XI/XI Plus들이 단행본의 주제분류를 위한 전문가시스템의 개발에 적합한지 여부를 다음의 다섯가지 기준에 의해 조사하였다. ①산출된 분류번호의 정확성, ②시스템의 응답시간, ③이용의 용이성, ④지식베이스(분류표)의 수정 가능성, ⑤시스템의 정보요구와 결론에 대한 이유를 설명하는 기능 등이다. 그리고 분류지식베이스는 모두 DDC 19판의 응용과학분야중 농학(630)을 택하여 구성하였다. 분석결과 이러한 셸들을 이용하면 시스템개발 시간을 절약할 수 있는 장점이 있지만 여러가지 제한점이 많아 오히려 Prolog나 LISP같은 언어를 이용하여 처음부터 분류전문가시스템을 개발

18) Sharif, Carolyn A. *Developing an Expert System for Classification of Books Using Micro-Based Expert System Shells.*(British Library Research Paper 32). Yorkshire : British Library Research and Development Department, 1988.

하는 것이 바람직하다고 주장하면서 분류의 전반적인 과정을 보다 철저하게 분석하고, 지식표현에 대한 새로운 방안을 모색할 것을 제안하고 있다.

### (1) Expert-Ease

Expert-Ease는 광범위한 분야에 적용될 수 있는 전문가시스템 작성도구라고 할 수 있겠는데 이는 입력된 사례의 집합으로부터 적당한 규칙들을 추론하고 이 규칙들을 결정나무(decision tree)로 표현하는 셸이다. 이 셸을 이용하여 DDC 19판의 10개 주류(main class)중 응용과학분야의 농학을 대상으로 분류전문가시스템을 구축하였는데 각 사례는 예를들면 다음과 같은 IF-THEN형식으로 표현되었다.

IF 주류가 응용과학이고 제1세목이 농업이며 제2세목이 곤충재배(양봉 및 양잠)이면 THEN 분류번호는 638이다.

이 분류전문가시스템은 맨처음 “분류하고자 하는 단행본의 주제가 10개의 세목중에서 어느 것에 해당되는지 선택을 하시오”라는 내용의 메뉴를 이용자에게 제공한다. 여기서 하나의 항목(제1세목)을 선택하면 그 다음에는 선택된 항목의 세부항목(제2세목)들을 열거한 메뉴가 제공되어 계속 해당되는 항목을 선택하게 되며, 이러한 과정은 적절한 분류번호가 산출될 때까지 계속된다.

Expert-Ease로 분류전문가시스템을 구축할 경우 시스템구축자와 최종이용자가 사용하기 쉽고, 간단한 모형을 단시간내에 구성할 수 있으며, 또한 시스템구축자가 전문가의 지식을 규칙형태로 얻지 않아도 되는 이점들이 있긴 하지만 몇 가지 제한점을 갖고 있다. 첫째 제한점으로는 분류하고자 하는 단행본의 표제, 목차 등에서 출현하는 단어들을 입력하여 문헌을 분류하는 대신 메뉴방식으로 주제용어를 선택해 가기 때문에 어느 정도 경험이 있는 분류전문가라야 사용할 수 있다는 점이다. 다음은 방대한 분류표의 내용가운데 극히 일부분만을 입력하여 처리할 수 있기 때문에 실제 분류전문가시스템으로 활용하기에는 제한점이 많다는 점이다.



## (2) ESP/Advisor

이는 전문적인 상담업무에 적합한 전문가시스템 셀인데 지식베이스는 "Sections"와 "Paragraphs"라고 하는 수많은 부분으로 이루어져 있다. 여기서 분류표의 내용을 지식베이스로 구축할 때 지식표현언어(KRL)를 사용해야 하는데 KRL은 다소 까다로운 문법구조를 갖고 있다. 따라서 KRL에 익숙해지기 까지 다소 시간이 걸리므로 시스템구축에 드는 시간이 Expert-Ease를 이용할 경우보다는 더 많이 걸리게 된다. 그렇지만 지식공학자가 Easywriter나 WordPerfect와 같은 워드프로세서를 이용하여 지식베이스의 소스파일을 작성하면 이 파일들은 ESP/Advisor KRL로 바꿀 수가 있다.

구축된 모형은 메뉴방식으로 구성되어 있으며, Expert-Ease를 이용한 경우에서 처럼 이용자는 분류할 문헌의 주제를 결정하고 적절한 분류번호가 산출될 때까지 메뉴로 부터 주류와 세목들을 선택한다. ESP/Advisor는 분류전문가시스템의 구축도구로서 Expert-Ease보다는 다소 성능이 나은 편이지만 여전히 많은 제한점들을 가지고 있어서 대규모 분류전문가시스템으로 활용하기에는 부적합하다.

## (3) Expertech XI/XI Plus

XI는 인간의 전문지식과 기술을 컴퓨터 프로그램으로 표현하기 위한 장치의 집합과 함께 영어와 유사한 언어로 구성된 전문가시스템 셀이다. 지식베이스를 구축하기 위해서 이용된 장치의 집합은 "Toolkit"라고 불리운다. 지식베이스는 두가지 방법 즉, 메뉴모드나 다이알로그모드에 의해 접근되어 진다. XI의 추론은 전진추론과 후진추론에 의존하며, 추론과정 및 논증의 정당성을 설명할 수 있는 기능 등 다양한 기능들을 갖고 있는 전문가시스템 셀이나 시스템의 응답시간이 매우 느린 것이 문제점으로 지적되고 있다. XI Plus는 시스템의 응답시간을 좀 더 빠르게 하는 등 XI의 기능을 수정보완한 셀이다. XI/XI Plus는 앞에서 언급한 두개의 셀보다도 다양한 기능을 더 많이 갖고 있기는 하나 분류표에 있는 모든 용어들을 포함하기에는 너무 많은 규칙이 요구된다. 또한 100개 미만

의 규칙들로 구성된 소규모 지식베이스를 이용하여 작동할 때에도 응답시간이 너무 느리다는 이유 등 때문에 분류전문가시스템으로 이용하기에는 적합하지 않은 셀들이다.

#### 7) INIS 데이터베이스 분류전문가시스템

1990년 최광<sup>19)</sup>은 그의 석사학위논문에서 INIS 데이터베이스의 분류표인 INIS Subject Categories and Scope Descriptions 의 플라즈마 연구부분인 A14.1대를 대상으로 분류표지식베이스와 분류기호 부여규칙을 이용한 자동분류시스템을 실험적으로 구축하였다.

이 시스템은 분류표 지식베이스, 분류표 지식베이스 색인파일, 분류규칙베이스, 분류표 지식베이스 탐색알고리즘으로 구성되며, 분류 대상문헌에서 분류자가 추출한 주제개념들을 입력받아 해당 분류기호를 출력하게 된다. 분류표 지식베이스에는 각각의 분류기호에 해당하는 지식이 축적되어 있는데 이는 INIS 디소러스와 플라즈마 연구분야의 전문가들로부터 수집하여 프레임구조로 구축되었다. 그리고 분류규칙베이스에는 적절한 분류기호를 부여하기 위한 추론기제가 들어 있으며, 이 규칙베이스의 통제를 받는 지식베이스 탐색알고리즘은 각각의 주제개념에 대해 지식베이스 색인파일과 분류표 지식베이스를 이용한 치환탐색기법을 이용하여 해당 분류기호를 부여한다. 치환탐색기법은 원하는 결과에 도달할 때까지 탐색된 결과를 탐색항목으로 치환하여 반복적으로 탐색하는 기법을 말한다. 이렇게 부여된 각각의 주제개념들에 대한 분류기호들을 분류규칙베이스에서 종합하여 그 문헌에 대한 적절한 분류기호를 부여하거나 상위분류기호를 부여하게 된다.

#### 8) AutoBC시스템

1992년 이경호<sup>20)</sup>는 그의 박사학위논문에서 콜론분류법(CC)의 패싱방식과 기

19) 최광, "분류표지식베이스를 이용한 자동분류에 관한 연구," 석사학위논문, 연세대학교 대학원, 1990.8.

20) 이경호, "콜론분류법에 바탕한 자동분류시스템의 개발에 관한 연구 ; 농학 및 의학 전문 도서관을 사례로," 박사학위논문, 성균관대학교 대학원, 1992.8.

호법을 약간 수정적용하여 전문도서관에서 실용화할 수 있는 자동분류(AutoBC) 시스템을 개발했는데 이는 분류자동화가 가능한 분류데이터베이스를 설계, 구축하고 도서의 표제나 키워드를 컴퓨터에 입력하므로써 주제의 자동적인 인지는 물론 표제속에 있는 키워드를 CC의 기호조합방식으로 처리하여 분류기호를 자동적으로 생성시킬 수 있는 시스템이다.

이 AutoBC시스템은 도서의 표제나 키워드를 입력하는 시스템, 입력된 데이터를 근거로 AutoBC 분류데이터베이스에서 키워드를 탐색하는 시스템, 탐색된 키워드를 주제분야별 출현빈도에 따라 주제를 인지하는 시스템, 주제인지 후 주제분야별 분류기호 조합원리에 따라 분류기호를 생성하는 시스템 및 분류데이터베이스의 갱신시스템의 5개 하부시스템으로 구성되어 있다. 여기서 가장 핵심되는 부분은 분류데이터베이스라고 할 수 있겠는데 이는 원통형의 원리에 의거하여 컴퓨터가 입력된 용어에서 특정주제의 인식과 더불어 분류기호를 CC의 규칙에 따라 조합할 수 있도록 되어 있다. 결국 이 시스템의 성패는 다른시스템과 마찬가지로 분류데이터베이스(지식베이스)내에 그 주제분야의 지식내용이 얼마나 잘 조직되느냐에 달려 있다고 하겠다.

지금까지 소개한 사례외에도 1989년 이용민<sup>21)</sup>은 DDC 20판 전산과학분야를 대상으로 분류표 지식베이스를 구축하여 분류표 지식베이스와 문헌검색의 결합 시스템을 실험적으로 개발하였으며, Gilroy<sup>22)</sup>는 DDC를 대상으로 사전 지식이 거의 없는 사람들도 이용하기 쉬운 Crystal이라는 셸을 이용하여 분류전문가시스템을 개발하였지만 이는 앞서 언급한 사례들과 거의 유사하다.

21) 이용민, "온라인목록 주제접근점 확장을 위한 분류표지식베이스설계에 관한 연구," 석사학위논문, 연세대학교 대학원, 1990.2.

22) Joyce Gilroy, "An Expert System for Classification," *Its News*, No.22 (August 1990), pp.39-40.

## IV. 결 론

이상에서 살펴본 바와 같이 분류전문가시스템도 다른 일반적인 전문가시스템과 기본구조는 별로 다를 바가 없다. 즉, 지식베이스를 정점으로 추론엔진, 지식 획득시스템, 이용자 인터페이스, 설명기능 등이 필요하며, 지식내용과 구조화방법이 다른 것이다. 그리고 전문가시스템의 사례분석에서 알 수 있듯이 이들은 대부분 지식베이스내에 들어갈 분류지식의 축적방법, 이를 탐색하여 분류기호를 부여하기 위한 분류규칙, 어휘통제의 방법, 분류자의 정보처리 행태에 관한 연구와 사용중인 분류표를 부분적으로 시스템화하여 시험한 것들이다. 시스템화하여 사용한 것으로는 주로 국제십진분류표, 일본십진분류표, 듀이십진분류표, 그리고 국제특허분류표를 비롯한 기타 특수분류표 등을 대상으로 분류표의 검색기능과 분류기호 조합기능을 적절히 결합, 개발한 것으로 시스템과 이용자 사이의 인터페이스가 강조되고 있다.

결국 분류전문가시스템을 개발하기 위해서는 전문가시스템 쉘을 이용하는 것보다 Prolog나 LISP와 같은 언어를 이용하여 새로이 시스템을 구축하는 것이 바람직하며, 기본적으로 다음의 요건들이 충족되어야 할 것이다.

첫째, 문헌의 주제개념이 정확하게 추출되어야 한다.

둘째, 특정 주제분야만이 아니라 분류에 필요한 모든 주제분야의 지식이 지식베이스화 되어야 한다.

셋째, 지식베이스는 분류지식의 특성을 잘 나타낼 수 있도록 조직되어야 한다.

넷째, 분류기호 부여규칙에 대한 지식베이스가 정확하게 이루어져야 한다.

다섯째, 분류기호가 즉각적으로 출력될 수 있어야 한다.

## 참 고 문 헌

- 이경호. “콜론분류법에 바탕한 자동분류시스템의 개발에 관한 연구;농학 및 의학 전문 도서관을 사례로,” 박사학위논문, 성균관대학교 대학원, 1992.8.
- 이용민. “온라인목록 주제접근점 확장을 위한 분류표지식베이스설계에 관한 연구,” 석사학위논문, 연세대학교 대학원, 1990.2.
- 정영미. “도서관·정보학에서의 인공지능의 응용에 관한 고찰,” 圖書館學, 제14집 (1987), pp.67-92.
- 최 광. “분류표지식베이스를 이용한 자동분류에 관한 연구,” 석사학위논문, 연세대학교 대학원, 1990.8.
- Aluri,R. and Riggs,D.E. “도서관에서의 전문가시스템 응용,” 김현희 역, 도서관, Vol.45, No.5(1990, 9·10), pp.27-45 / No.6(1990, 11·12), pp.12-29.
- 細野公男, 田村俊作, 原田隆史. “圖書分類エキスパートシステム,” 情報學基礎, 8-4 (1988), pp.1-7.
- 石川徹也, 篠原まさ子. “日本十進分類法(NDC)を對象とする分類番號附與支援システム,” 情報科學技術研究集會發表論文集, 第22回(1985), pp.201-209.
- Bobrow, D.G. and Mittal, S. “Expert Systems : Perils and Promise,” *Communications of the ACM*, Vol.29, No.9(1986), pp.880-894.
- Clarke, A and Cronin,B. “Expert Systems and Library/Information Work,” *Journal of Librarianship*, Vol.15, No.4(1983), pp.277-292.
- Endres-Niggemeyer, Brigitte and Schmidt,Bettina. “Knowledge Based Classification Systems ; Basic Issues,a Toy System and Further Prospects,” *International Classification*, Vol.16, No.3(1989), pp.146-156.

- Ford, Nigel. *Expert Systems and Artificial Intelligence ; an Information Manager's Guide*. London : LA, 1991.
- Gibb, Forbes ed. *Expert Systems in Libraries*. London : Taylor Graham, 1986.
- Gilroy, Joyce. "An Expert System for Classification," *Its News*, No.22 (August 1990), pp.39-40.
- Hayes-Roth, Fredrick etc. ed. *Building Expert Systems*. London : Addison-Wesley, 1983.
- Ishikawa T. "The Man-Machine Interface Aspect of an Automatic Classification Numbering System in a Computerized Library System," *Journal of Information Processing*, Vol.11, No.3(1988), pp.199-205.
- Nakamura Y. and Ishikawa T. "Expert System for Automatic UDC Numbering Assignment," *Proceedings of the 44th FID Conference and Congress part I (He Isingki, 1988)*, pp.274-283.
- Rolston, D.W. *Principles of Artificial Intelligences and Expert Systems Development*. New York : McGraw-Hill, 1986.
- Sharif, Carolyn A. *Developing an Expert System for Classification of Books Using Micro-Based Expert System Shells*. (British Library Research Paper 32). Yorkshire : British Library Research and Development Department, 1988.
- Valkonen, Pekka and Nykanen, Olli. "An Expert System for Patent Classification," *World Patent Information*, Vol.13, No.3(1991), pp.143-148.
- Waterman, Donald A. *A Guide to Expert Systems*. London : Addison-Wesley , 1986.

## A Study on the Expert System for Classification of Books

Kim Jeong-hyen\*

## &lt;Abstract&gt;

This study is an attempt to provide some helpful data for the design and the implementation of the expert system for the book-classification based on the analysis of various cases of the classification-expert system models.

Following the introduction, the concepts and some features of an expert system were overviewed in the second chapter, on the basis of which the following concrete cases were introduced and analyzed in the third chapter : ①ACN System for NC, ②Expert System for NDC, ③Expert System for UDC, ④Herba Medica System, ⑤Expert System for IPC, ⑥Stratcyclode Project, ⑦Expert System for Classification of INIS Database, ⑧AutoBC System, and etc.

In the conclusion, for the development of the classification-expert system, it was turned out that constructing a new system by using an AI language such as Prolog or LISP is more desirable than employing any one of expert system shells. Together it is necessary for the following requirements to be met : ①The subject concept of a document elicited should be accurate. ②Not only a domain knowledge but also the knowledge covering all the subjects should be represented in the knowledge-bases. ③The knowledge-bases should be organized in such a way that the characteristics of the knowledge about classification should be well defined. ④A rule-base consisting of accurate rules about classification should be made. ⑤It should be possible for classification code wanted to be generated immediately.

---

\* Keimyung University, Lecture.